

User guide

Executing recognizers under ELAN

Christopher Rosenthal

Przemyslaw Lenkiewicz

Diana Ransgaard Sørensen

Content

1. Install ELAN 4.1.2.....	4
2. Install GuiSkin	4
3. ELAN extensions	4
4. Using the ELAN audio recognizers.....	5
5. IAIS 02: Standard homogeneous segmentation	6
6. IAIS 03: Fine audio segmentation	8
7. IAIS 04: Speech/ Non-Speech detection based on pre-trained acoustic models.....	9
8. IAIS 06: Speaker diarization	10
9. Skin Color Estimation with GuiSkin	11
10. Using the hands and head tracking recognizer	12
Appendix I	15

List of figures

Figure 1_ Extraction of extensions.....	4
Figure 2_ Open new audio file	5
Figure 3_ IAIS 02: Standard audio segmentation	6
Figure 4_ ELAN automatically saves the output	7
Figure 5_ Fine audio section.....	8
Figure 6_ IAIS 04: Speech/ Non-Speech recognizer.....	9
Figure 7_ Speaker diarization	10
Figure 8_ Skin color estimation.....	11
Figure 9_ Open (new) video file.....	12
Figure 10_ Video Recognizer and Parameters.....	13
Figure 11_ Create tiers from segment.....	14
Figure 12_ Example of four tiers with annotation	14

1. Install ELAN 4.1.2.

The ELAN annotation tool allows accessing the recognizers that are running on the Max Planck servers. In order to use this functionality, first [download ELAN](#).

Note: ELAN can be placed on any drive. For example:

C:\Program Files (x86)\ELAN 4.1.2

2. Install the Skin Color Estimation Application

Necessary for optimal configuration of the Hands and Head Tracking Recognizer. It can be downloaded from the following location. For instructions on using please see section 9 on page 11.

3. ELAN extensions

Download the [extensions files](#) and extract them (Figure 1_ Extraction of extensions) to your ELAN/extensions folder.

If ELAN is installed at the C-drive, then place the downloaded extensions at:

C:\Program Files (x86)\ELAN 4.1.2\extensions

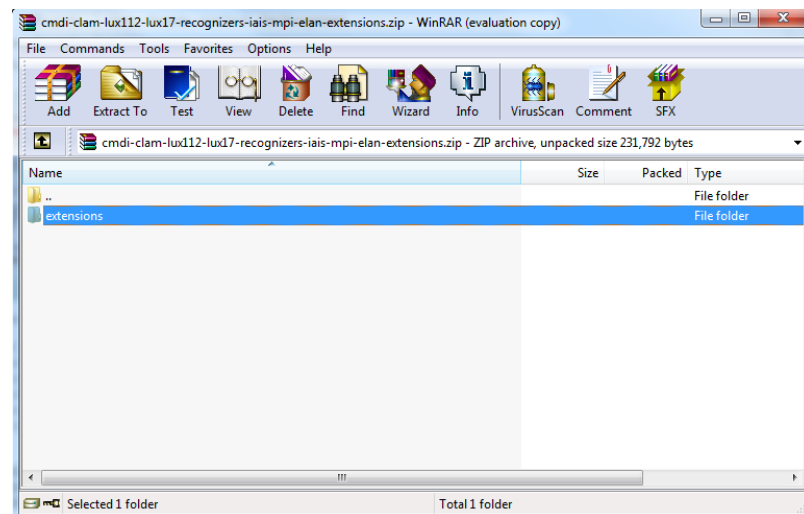


Figure 1_ Extraction of extensions

4. Using the ELAN audio recognizers

To start using the audio recognizers, open ELAN, go to -> FILE -> NEW. **Open the desired** audio file by **moving it into the Selected Files** menu (Figure 2_ Open new audio file). Then click OK.

Note: The files can be placed on any drive. Go to the location (either a folder or the desktop), and create a new folder called e.g. **Recognizers**. The folder must be accessible with write permissions for the recognizer. Select the audio files desired and copy/paste them to the newly created location.

Note: You cannot process a video file with audio recognizers directly. You need to have the video file and also a separate WAV file with the audio track. It is a common practice to obtain both files when digitizing the recordings at MPI. When processing such video recording, make sure to include both files (audio and video) in the **Selected Files** window.

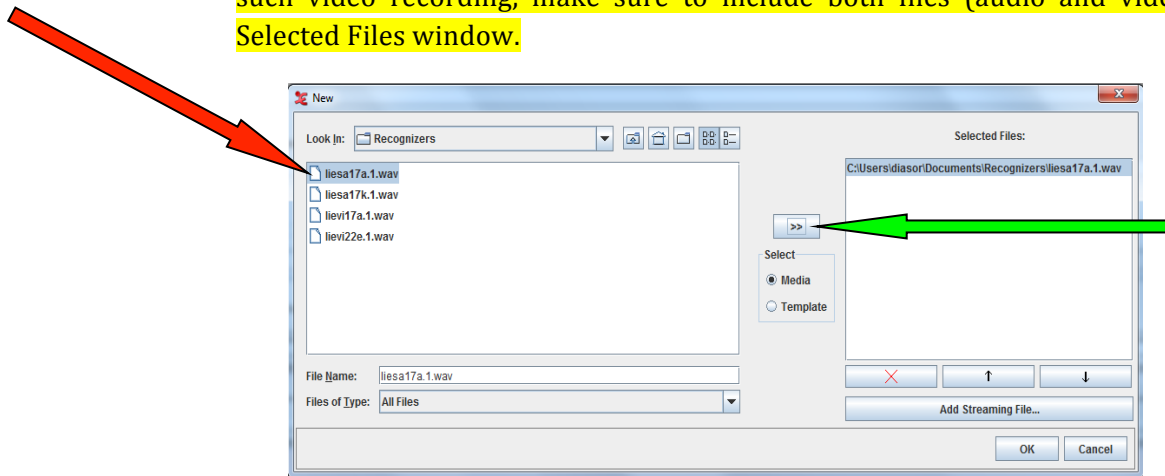


Figure 2_ Open new audio file

To let ELAN create audio segmentations within an audio file the program offers two recognizers:

1. **IAIS 02: Standard homogeneous segmentation** that splits audio on significant changes (e.g. new speaker, music).
2. **IAIS 03: Fine audio segmentation** for splitting audio into utterance level segments.

Both of them can be used for audio segmentation, however, the **IAIS 03: Fine audio segmentation** gives the user more control to fine-tune the results.

5. IAIS 02: Standard homogeneous segmentation

Choose the **IAIS 02: Standard homogeneous segmentation** recognizer from the Recognizer dropdown list (Figure 3_ IAIS 02: Standard audio segmentation). **Define the output file**, where the results of the segmentation will be stored. The file has to be located in a folder, which is accessible with write permissions for the recognizer, so e.g. the **Recognizers** folder that has been created in **section 4** (Using the ELAN audio recognizers). The name of the file needs to end with ".xml". Try to give this file a meaningful name (as the file will later be reused), like:

OUT [xml] Tier holding the standard segmentation (e.g.):

C:\Users\diasor\Documents\Recognizers\output.audio.standard.segmentation.xml

Click on **START** and let ELAN define segments within your audio file.

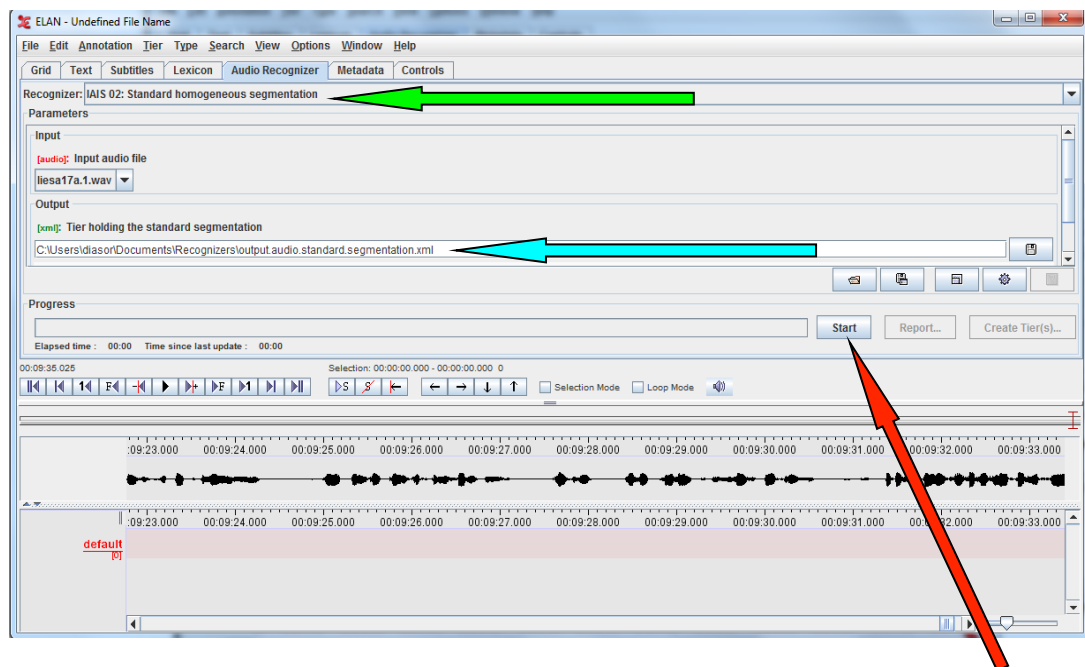


Figure 3_ IAIS 02: Standard audio segmentation

If the output file is **not defined**, ELAN **automatically saves the output** in the **Recognizers** folder created in **section 4** (Using the ELAN audio recognizers).

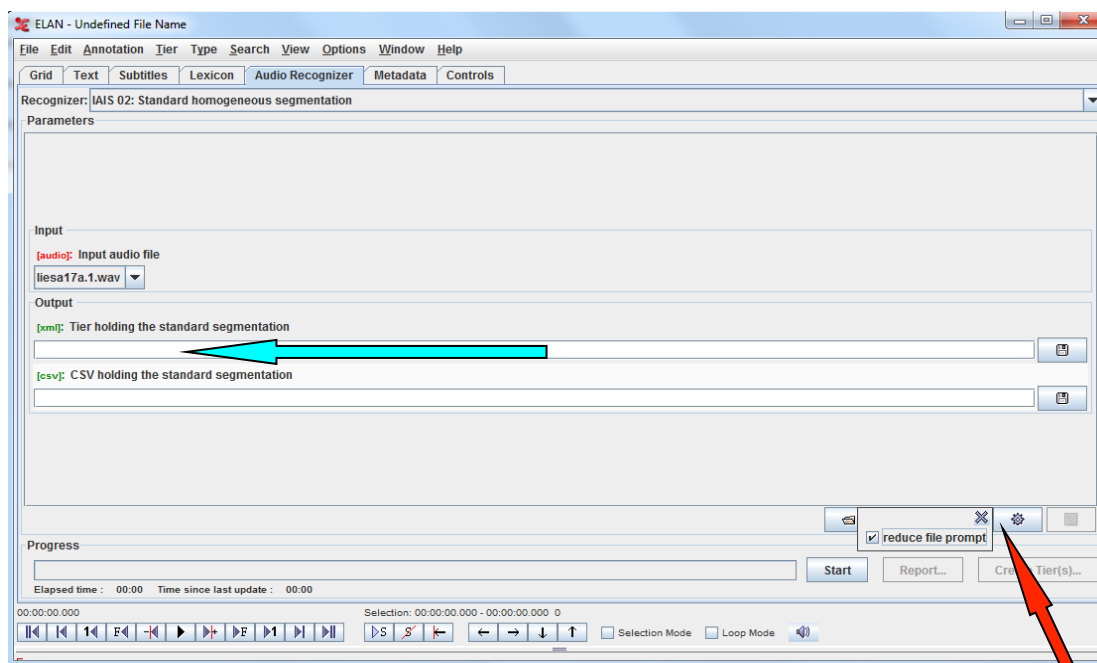


Figure 4_ ELAN automatically saves the output

6. IAIS 03: Fine audio segmentation

The IAIS 03: Fine audio segmentation works in a similar way to the IAIS 02 recognizer. The difference is that it offers the user the chance to modify a parameter, which controls the sensitivity of the segmentation process.

Choose the fine audio segmentation (Figure 5_ Fine audio section) and Define the output file, where the results of the segmentation will be stored.

OUT [xml] Tier holding the fine segmentation (e.g.):

C:\Users\diasor\Documents\Recognizers\output.audio.fine.segmentation.xml

Use the slider to tune the sensitivity of the recognizer and the size of the resulting segments.

Choose whether the recognizer should perform merging of the results. This step will merge the resulting segments that are neighbors and have high similarity.

Click on **START** and let ELAN define segments within your audio file.

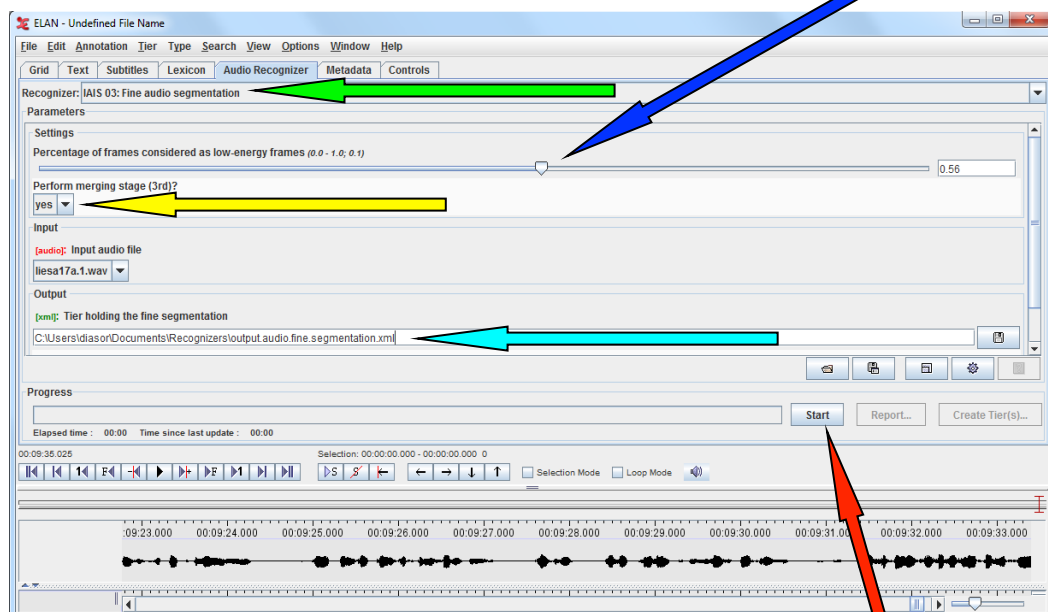


Figure 5_ Fine audio section

7. IAIS 04: Speech/ Non-Speech detection based on pre-trained acoustic models

To detect what parts of the recordings that contains human speech, change the recognizer to **IAIS 04: Speech/ Non-Speech detection based on pre-trained acoustic models** (see Figure 6_ IAIS 04: Speech/ Non-Speech recognizerFigure 6_ IAIS 04: Speech/ Non-Speech). Within the parameter section fill in the input file name (tick file), which contains the segmentation information you have created earlier, and output file name, which will hold the speech/no speech results. For example:

IN [xml]: Result of the segmentation with optional manual labels for training (e.g.):

C:\Users\diasor\Documents\Recognizers\output.audio.fine.segmentation.xml

OUT [xml]: Tier holding the speech/non-speech segmentation (e.g.):

C:\Users\diasor\Documents\Recognizers\output.audio.speech.xml

Another option within the parameter section is to choose the tier – default (tick tier). It allows you to choose a tier you previously created with another recognizer.

Then click on **START**.

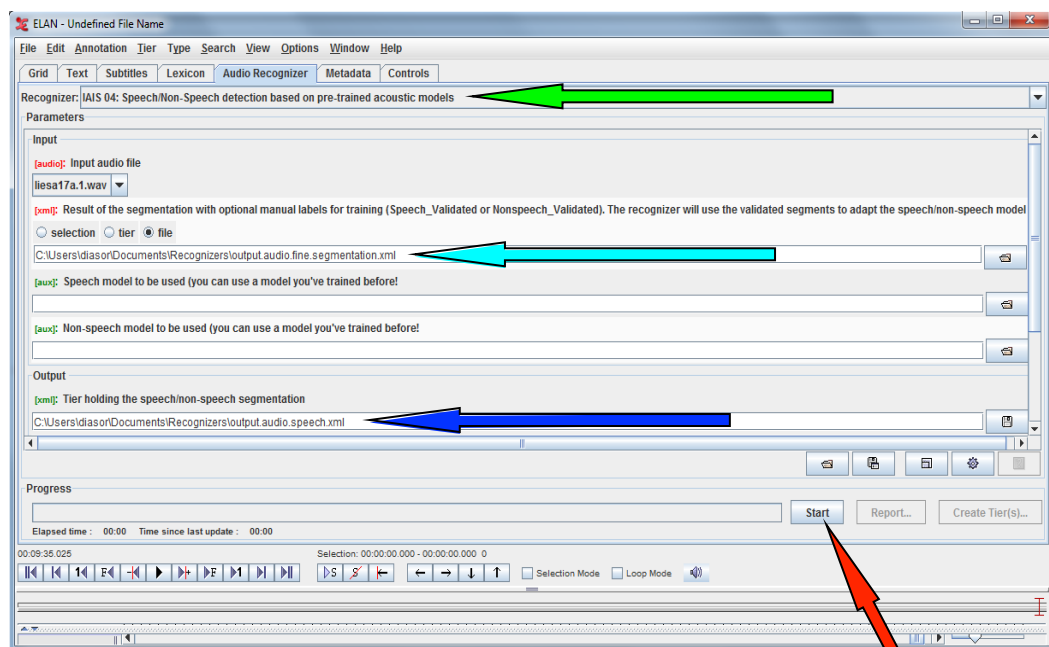


Figure 6_ IAIS 04: Speech/ Non-Speech recognizer

8. IAIS 06: Speaker diarization

The speaker diarization recognizer will try to assign each segment of the recording to respective speaker. So it should detect the number of speakers in the recording and detect who is speaking when.

Change the recognizer to **IAIS 06: Speaker diarization** (Figure 7_ Speaker diarization). Within the parameter section fill in the input file name (tick file), which contains the speech/non-speech information you have created earlier, and output file name, which will hold the results of speaker diarization. For example:

IN [xml]: Result of one of the speech/non-speech recognizers (e.g.):

C:\Users\diasor\Documents\Recognizers\output.audio.speech.xml

OUT [xml]: Tier holding the speaker diarization (e.g.):

C:\Users\diasor\Documents\Recognizers\output.audio.speaker.xml

Another option within the parameter section is to choose the tier – default (tick tier). It allows you to choose a tier you previously created with another recognizer.

Then click on **START**.

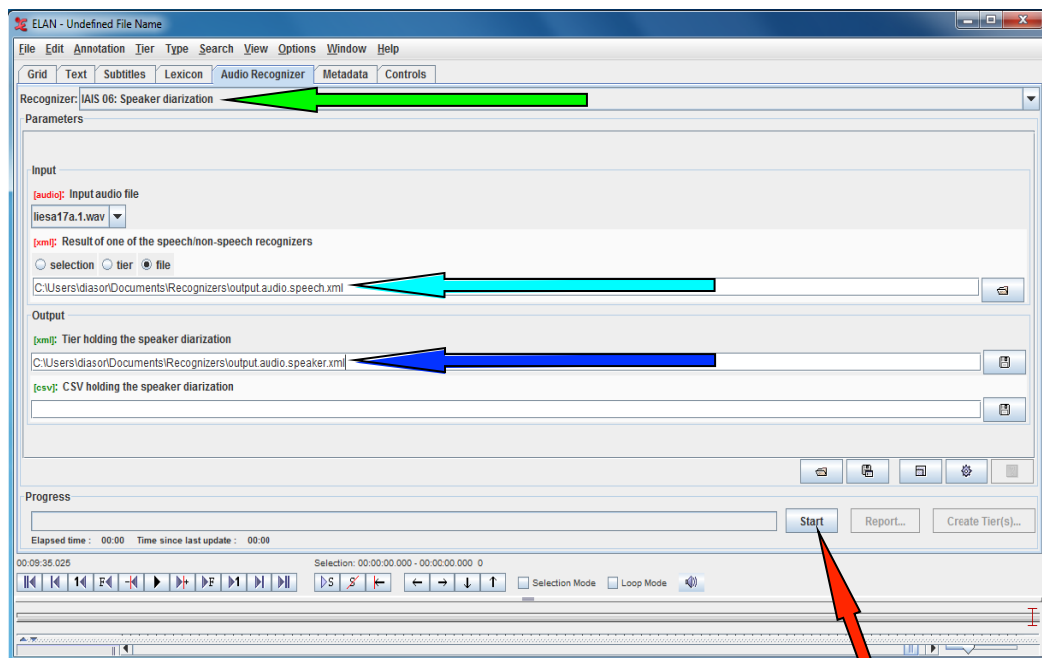


Figure 7_ Speaker diarization

9. Skin Color Estimation with GuiSkin

The Skin Color Estimation App is a graphical way to estimate skin color in a video by manually selecting the best intervals in the YUV domain. This step has to be done before the video file can be successfully processed by Hands and Head tracking recognizers.

- 1) Open GuiSkin and import a video (-> FILE -> LOAD VIDEO)
- 2) Use the color intervals to mark all skin with "blue ink" and make sure no background is marked

Advise: GuiSkin has six sliders: Y for the brightness, U and V for the color components. First three sliders represent the mid-point of the selected range; the last three sliders represent the size of the selected range. U usually ranges between 80 and 130, V between 125 and 175.

When estimating the color, you will notice this step involves trial and error. Our advice is to start with the second slider and observe the changes in the image. When a reasonable amount of skin color is highlighted, use the 5th slider to adjust the range for this value. The two remaining sliders can serve to limit the amount of non-skin pixels highlighted. Sometimes it cannot be avoided that some parts in the background will be marked as well. Priority should always be that all skin is marked. Non-moving background can be cancelled out by the recognizers later on.

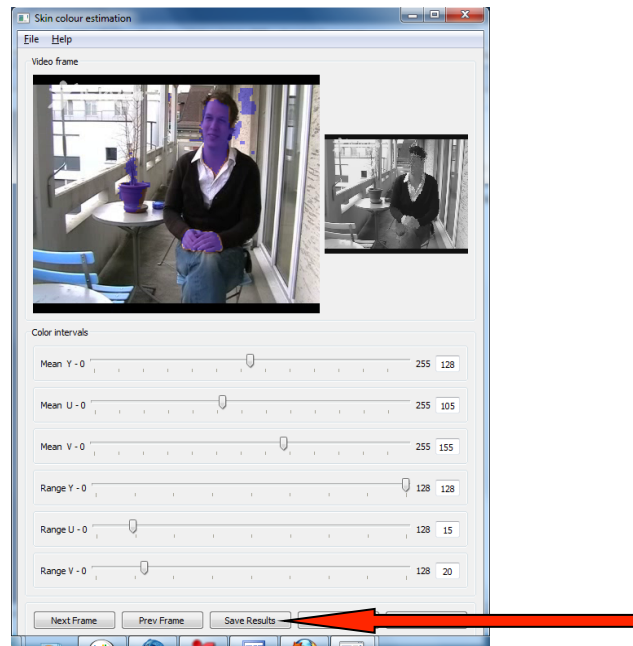


Figure 8_Skin color estimation

Once done calibrating, press SAVE RESULTS and save the given XML file in your **work folder from section 4: Using the ELAN audio recognizers** (or save it anywhere and then copy or move to your work folder).

10. Using the hands and head tracking recognizer

Open ELAN -> FILE -> NEW and move your video into the SELECTED FILES window and press OK.

Advise: for information about decent resolution, video quality and lighting condition, read: Appendix I– Guidelines for video capture.

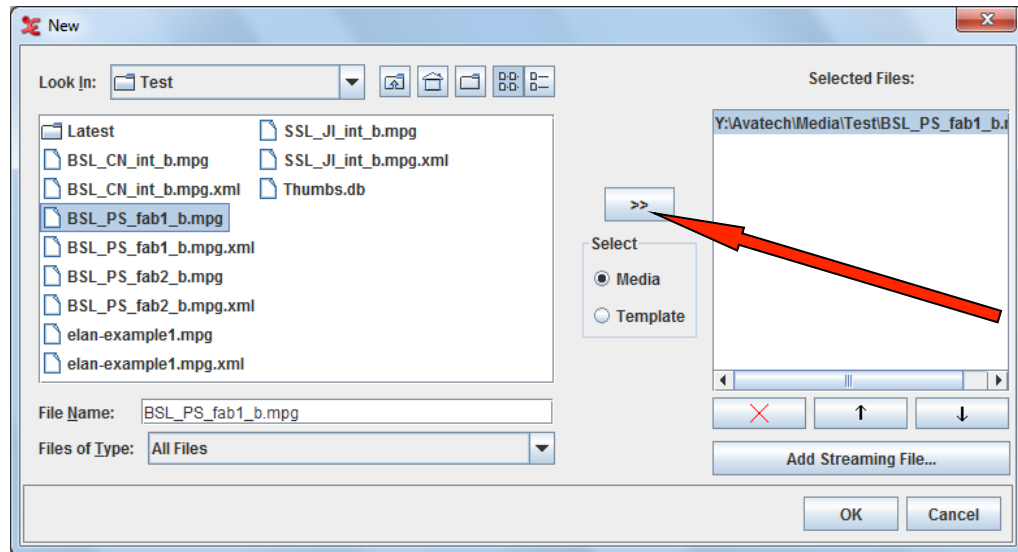


Figure 9_ Open (new) video file

To your upper right, click on **VIDEO RECOGNIZER** (Figure 10_ Video Recognizer and Parameters) and choose Tracks motion of hands and head.

Under **Parameters** you will see four empty text fields:

- 1) **IN [csv]: xml containing result of segmentation1 [XML_filename]**
Here you need to import the XML file you have created with GuiSkin in previous section.
- 2) **OUT [aux] xml updated with hands movement information**
This file holds the resulting tiers and annotation segments, describing which body parts are moving.
Fill in: X:\recognizers\output.mpg.xml
- 3) **OUT [csv] csv files with hand/head frame by frame information**
This file holds the position of the tracked body parts for each frame of the video using X and Y coordinates.
Fill in: X:\recognizers\output.csv
- 4) **OUT [aux]: video with overlayed hands/heads tracked**
Here you can save the video with ellipses over the tracked body parts.
Fill in: X:\recognizers\output.avi
Once filled in, press START

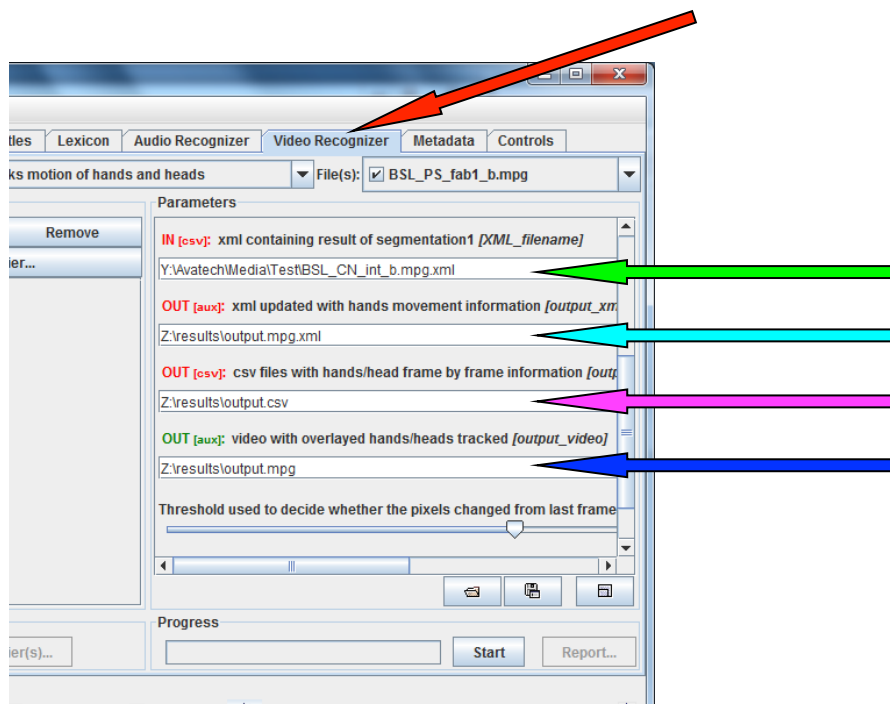


Figure 10_ Video Recognizer and Parameters

This recognizer in current version is somewhat unstable and it can either crash or endlessly keep you waiting for the results.

The way of checking if it's still working is to open your work folder and see if the files you have just created above are changing their size. If they're growing, the recognizer is still working. If not, it is finished. In such case press CANCEL once or twice (this is a bug and it should be fixed soon).

Once the program is finished, you have to import the tiers manually, one at a time.

Press FILE -> IMPORT -> IMPORT TIERS FROM RECOGNIZERS

Open output.mpg.xml (which you can find in your RECOGNIZERS folder) and import the tiers one at a time (i.e. by selecting just one tier at a time and clicking CREATE for each of them) (Figure 11_ Create tiers from segment).

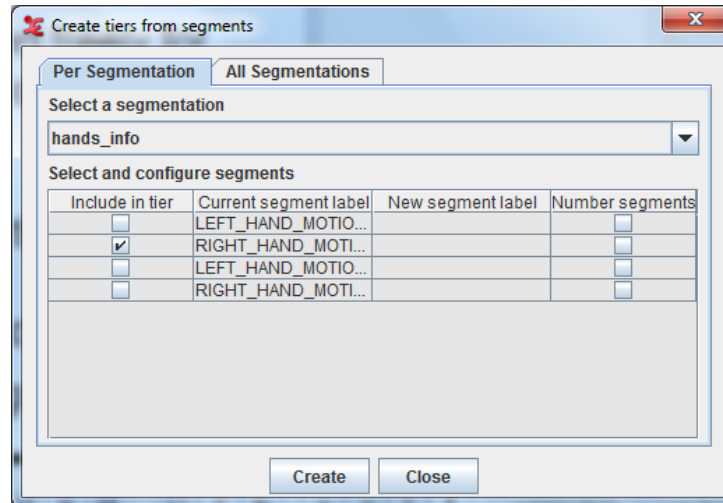


Figure 11_ Create tiers from segment

If everything has worked out ELAN should have created four tiers with annotation information (See example in Figure 12_ Example of four tiers with annotation).

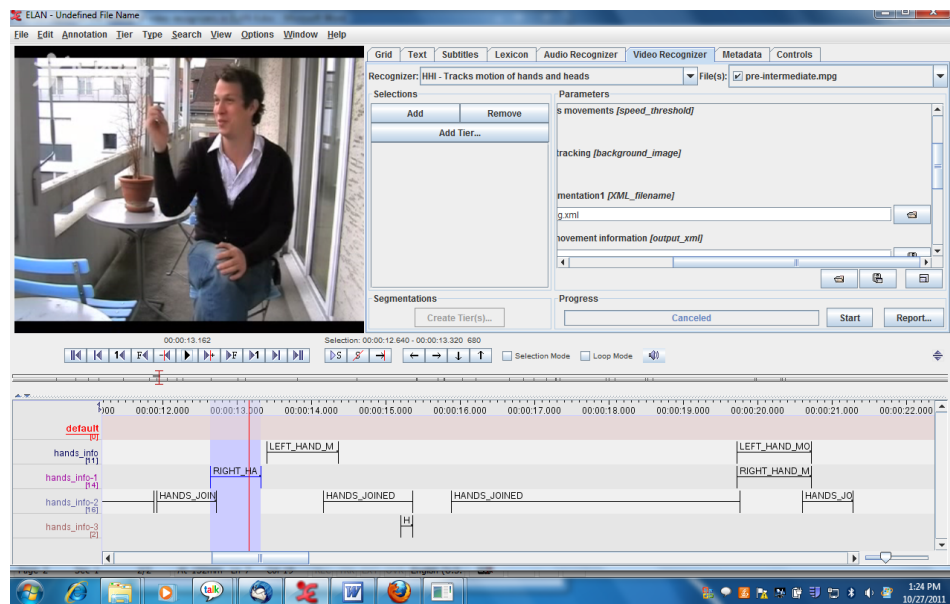


Figure 12_ Example of four tiers with annotation

Appendix I

Guidelines for video capture

- Decent resolution (at least standard definition, 720 X 576 pixels)
- Decent video quality (the higher the bitrate, the better)
- Uniform lighting condition, neither too dark nor too bright

Bad examples



Good examples



- The color of the clothes should be different than the color of the skin. Same for the background, if it is very close to the hands.

Bad examples



Good example



- No more than two persons in the scene.
- Fixed camera.

Other settings that are not mandatory, but that potentially yield better automatic annotations:

- People should face the camera.
- People should be close to the camera.

Bad example



Good example



- If there are two persons, they should be at the same distance from the camera.
- The tracking is easier if the person wears long-sleeves clothes instead of short-sleeves.
- Background removal works only with static or almost-static background (background removal is used only if the color of the objects in the background is similar to skin color, though).